DOCUMENT RESUME

ED 259 024 TM 850 396

AUTHOR Jones, Robert F.; Vanyur, Suzanne
TITLE An Investigation of Gender-Related Test Bias for the Medical College Admission Test.
PUB DATE Apr 85
NOTE 13p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 1-4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Females; Grade Point Average; Grade Prediction; *Graduate Medical Education; Higher Education; Males; Multiple Regression Analysis; Predictive Validity; Science Tests; *Sex Differences; *Test Bias; Test Use
IDENTIFIERS *Medical College Admission Test

ABSTRACT

The Medical College Admission Test (MCAT) was examined for sex bias. Data for student samples from two midwestern state-supported medical schools were analyzed. The criteria included average course grades in each of the first two years of medical school. Tests comparing various regression models based on male and female subgroups generally failed to find significant differences in standard errors of estimate, regression slopes, or intercepts. The few exceptions to this overall finding centered on two regression models which included scores on the physics subtest, and to a lesser extent, the chemistry subtest. In conclusion, this study failed to find great support for the assumption of widespread bias in predicting the medical school performance of females from MCAT scores, as long as undergraduate grade point average and Skills Analysis: Reading scores were considered in the selective index. Caution was also advised in the weighting of physics scores. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

An Investigation of Gender-Related Test Bias for the Medical College Admission Test

ED259024

Abstract

This study examines the presence of gender-related test bias in the Medical College Admission Test (MCAT). Data for student samples from two midwestern state-supported medical schools were analyzed. The criteria include average course grades in each of the first two years of medical school. Tests comparing various regression models based on male and female subgroups generally failed to find significant differences in standard errors of estimate, regression slopes, or intercepts. The few exceptions to this overall finding centered on two regression models which included MCAT Chemistry and Physics scores. The authors discuss the implications of these findings for the use of MCAT scores in medical school admissions.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Jones, Robert F.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 850 396

AN INVESTIGATION OF GENDER-RELATED TEST BIAS FOR THE MEDICAL COLLEGE ADMISSION TEST

Robert F. Jones, Ph.D.
Association of American Medical Colleges

Suzanne Vanyur
Federal Bureau of Prisons

The purpose of this study is to investigate the presence of gender-related test bias in the Medical College Admission Test (MCAT). The study approaches the issue of test bias from a perspective and methodology proposed by Cleary (1968). She defined a test to be biased for members of a subgroup if, "in the prediction of a criterion for which the test was designed, consistent nonzero errors of predictions are made for members of the subgroup".

This approach has now been used extensively to study possible bias in Scholastic Aptitude Test (SAT) scores, American College Testing Program (ACT) scores, and Law School Admission Test (LSAT) scores. In summarizing these studies, Linn (1982) observed the tendency for the test scores of females to correlate more strongly with performance measures and consequently to predict the subsequent academic performance of females somewhat more accurately than males. In those cases where the accuracy of prediction was similar, test scores have tended to underpredict systematically the performance of females. No studies of this type have yet been performed for the MCAT, which underwent a major revision in 1977. The fact that women tend to achieve lower scores than men in five of the six MCAT areas of assessment underscores the relevance and importance of such a study for medical school admissions (AAMC, 1983).

Paper presented at the meeting of the National Council on Measurement in Education, Chicago, Illinois, April, 1985.

BEST COPY

Instrument

The MCAT is a national standardized examination used in the selection of students for medical school in the United States and Canada. The day-long test consists of four sections: Science Knowledge (SK), Science Problems (SP), Skills Analysis:Reading (SA:R), and Skills Analysis:Quantitative (SA:Q). Six scores are provided on a scale which ranges from 1 to 15. Discipline scores in biology (BIO), chemistry (CHEM), and physics (PHY) are computed from performance on knowledge and problems items within each discipline, contained in the first two sections. The Science Problems score is based on all problems items across disciplines. Skills Analysis:Reading and Skills Analysis:Quantitative Scores are based on items in each of their respective sections.

Method

The study was conducted with student samples from each of two midwestern, state-supported medical schools. At each school, data were combined for two classes who had completed the first two years of medical school, which consist primarily of lectures and labs in the basic medical sciences. Courses in these two years generally cover areas of anatomy, physiology, biochemistry, microbiology, pharmacology, pathology, and behavioral science. At each school, the criterion measures used were average course grades within each year (GPA1 and GPA2). The grades for both schools are reported on a 5-point scale, where 5 represents "A" or "superior," and 1 represents "F" or "unsatisfactory." For all students, data on the undergraduate college grade-point average (UGPA), standardized to a four-point scale, were available in addition to MCAT scores.

BEST COPY

The analysis and comparison of prediction systems for men and women based on regression models proceeded using a sequential testing approach advocated by Gulliksen and Wilks (1950). In this approach, standard errors of estimate from the regression models based on each group are first compared, followed by tests of the equality of regression slopes, and then tests of the equality of intercepts. Failure to reject the null hypothesis at any stage precludes subsequent tests. Determining the specific regression models to test presented a dilemma. The overlap in item scoring previously described introduces a spurious correlation between Science Problems scores and Biology, Chemistry, and Physics scores. A multiple regression analysis involving all six MCAT scores has, therefore, inevitable problems of multicollinearity, and results in a large number of variables in the model relative to sample size. Also, based on a discussion by Linn and Werts (1971), the authors felt it important to include undergraduate grade-point average (UGPA) in each model. The decision was made to test regression models for each of the five MCAT areas of assessment where females tend to achieve lower scores than males, with UGPA and SA:R included in each model. UGPA is generally included in indexes used by medical schools in the initial screening of applicants. Skills Analysis:Reading is the one MCAT area of assessment in which females generally perform better than males.

Results

Tables 1 and 2 show means, standard deviations, and intercorrelations for predictor and criterion variables for School A and School B, respectively. As expected, females in School A have lower scores in Biology, Chemistry, Physics, and Science Problems, with the largest difference occurring in Physics. Scores on Skills Analysis:Reading and undergraduate grade-point

BEST COPY

Table 1

Means, Standard Deviations, and Intercorrelations Among Predictor and Criterion Variables for Males and Females - School A

	Males (N=198)		Females (N=54)	
	Mean	S.D.	Mean	S.D.
UGPA	3.49	.35	3.57	.32
BIO	8.73	1.76	8.22	2.15
CHEM	8.63	1.79	7.70	2.08
PHY	8.85	1.94	7.65	1.85
SP	8.89	1.65	8.20	1.97
SA:R	9.01	1.75	9.33	1.95
SA:Q	8.68	1.90	8.74	2.08
GPA1	2.53	.58	2.55	.56
GPA2	2.55	.62	2.72	.62

Intercorrelations*

	UGPA	BIO	CHEM	PHY	SP	SA:R	SA:Q	GPA1	GPA2
UGPA		.29	.46	.33	.32	.18	.22	.49	.37
BIO	-.04		.53	.30	.66	.45	.43	.35	.29
CHEM	-.03	.26		.55	.77	.39	.60	.39	.32
PHY	.09	.21	.53		.68	.14	.45	.19	.23
SP	.00	.54	.69	.61		.41	.61	.32	.23
SA:R	.05	.28	.24	.19	.31		.63	.40	.20
SA:Q	.03	.32	.39	.44	.49	.52		.45	.37
GPA1	.29	.24	.30	.17	.29	.16	.19		.74
GPA2	.32	.22	.19	.11	.21	.15	.13	.80	

*correlations for females are above the diagonal; for males, below the diagonal

Table 2

Means, Standard Deviations, and Intecorrelations Among Predictor and Criterion Variables for Males and Females - School B

	Males (N=263)		Females (N=94)	
	Mean	S.D.	Mean	S.D.
UGPA	3.62	.27	3.68	.24
BIO	9.20	1.56	9.03	1.53
CHEM	9.54	1.71	9.10	1.65
PHY	9.56	1.72	8.19	1.30
SP	9.63	1.45	9.14	1.48
SA:R	9.10	1.57	9.55	1.56
SA:Q	9.52	1.55	9.10	1.57
GPA1	3.48	.60	3.58	.63
GPA2	3.54	.59	3.64	.56

Intercorrelations*

	UGPA	BIO	CHEM	PHY	SP	SA:R	SA:Q	GPA1	GPA2
UGPA		.01	.27	.17	.11	.17	.07	.40	.26
BIO	-.08		.43	.31	.62	.24	.22	.38	.42
CHEM	.10	.35		.51	.70	.27	.37	.52	.51
PHY	.02	.29	.52		.59	.25	.22	.36	.48
SP	-.05	.54	.71	.61		.26	.43	.43	.47
SA:R	-.05	.26	.14	.24	.23		.36	.06	.22
SA:Q	-.03	.18	.35	.32	.38	.36		.11	.25
GPA1	.40	.21	.23	.14	.18	.02	.05		.70
GPA 2	.33	.27	.16	.14	.17	.10	.08	.71	

*correlations for females are above the diagonal; for males, below the diagonal.

average tend to favor females. While grades in the first year of medical school for males and females are very similar, females tend to do slightly better than males in the second year. Correlations of MCAT scores and UGPA with the criterion measures tend to be stronger for females than males, but this may be partly a consequence of larger variances for females on some of these measures.

Data for School B show many of the same patterns. However, in this sample, males achieve higher scores on Skills Analysis:Quantitative, consistent with the national data. Also, females achieve higher criterion scores in both years of medical school, although the differences are not large. Correlations of MCAT scores with the two criterion measures are substantially larger in the female group.

Table 3 provides for School A the results of significance tests comparing five regression models for males and females, in terms of standard errors of estimate, slopes, and intercepts. Although there is a tendency toward differences in intercepts in models predicting GPA2 (in the direction of female performance being underpredicted by the male equation), none of the thirty tests performed show differences that are significant at the .05 level. The results of similar tests for School B, shown in Table 4 did produce significant differences for two of the models tested. Models which included Chemistry and Physics scores in the prediction of GPA2 show differences in slopes. Further analysis revealed that these overall differences were due specifically to differences in the regression coefficients associated with Chemistry ($F=7.23$, $p<.01$) and Physics ($F=8.99$, $p<.01$) scores, respectively. Differences in intercepts were also found in the model which includes Physics scores to predict GPA1. In this model the performance of females tends to be underpredicted systematically by the equation based on males. The magnitude

BEST COPY

Table 3

Summary of Significance Tests Comparing
Regression Models for Males and Females - School A

Regression Model	F - tests*		
	Equality of Errors of Estimate	Equality of Slopes	Equality of Intercepts
UGPA, SA:R, and BIO			
GPA1	1.27	.55	.02
GPA2	1.04	.10	2.64
UGPA, SA:R, and CHEM			
GPA1	1.22	.67	.10
GPA2	1.04	.10	3.26
UGPA, SA:R, and PHY			
GPA1	1.31	.84	.03
GPA2	1.00	.10	2.01
UGPA, SA:R, and SP			
GPA1	1.22	.89	.02
GPA2	1.04	.19	2.73
UGPA, SA:R, and SA:Q			
GPA1	1.36	.84	.14
GPA2	1.07	.68	1.49

*none of the differences summarized in this table is significant at the .05 level.

Table 4

Summary of Significance Tests Comparing
Regression Models for Males and Females - School B

Regression Model	F - tests: Equality of Errors of Estimate	Equality of Slopes	Equality of Intercepts
UGPA, SA:R, and BIO			
GPA1	1.00	.49	1.21
GPA2	1.21	.90	1.63
UGPA, SA:R, and CHEM			
GPA1	1.07	1.32	2.18
GPA2	1.36	3.02*	-
UGPA, SA:R, and PHY			
GPA1	1.00	1.09	4.50*
GPA2	1.31	3.44*	-
UGPA, SA:R, and SP			
GPA1	1.07	1.15	2.92
GPA2	1.26	2.18	2.34
UGPA, SA:R, and SA:Q			
GPA1	1.07	.15	.72
GPA2	1.12	.69	1.03

*$p < .05$

of this underprediction was further explored by comparing the predicted GPA1 for females at the mean using a combined male-female equation with the actual GPA1. This difference was on the order of one-fourth (.24) of a standard deviation.

Conclusion

Given the number of statistical tests performed, the few differences found obviously need to be interpreted with caution. Still two of these differences centered on Physics scores, where the largest gap in test performance between males and females occurs (AAMC, 1983). Moreover, the direction of these few differences is consistent with previous studies of other admissions tests referred to earlier.

While the performance gap between males and females in scientific and quantitative areas of assessment is well documented (Maccoby and Jacklin, 1975) and much discussed, the authors are unaware of any theoretical postulates explaining the disjuncture between test performance and the subsequent academic achievement of females. Wang and Wilson-Pessano (1982) have studied bias at the item level for the MCAT and concluded that science items with a secondary mathematical operation are systematically more difficult for females. In the science areas of assessment, Physics items are most likely to include a mathematical operation, Chemistry items next most likely, and Biology items least likely, an ordering that parallels the size of the mean performance differences on the test. Medical school examinations may include fewer of these computational problems, a fact that would provide some explanation for the underprediction, the differences in intercepts, but not the differences in slopes.

In general, this study fails to find great support for the assumption of

BEST COPY

a widespread bias in predicting the medical school performance of females from MCAT scores as long as undergraduate grade-point average and Skills Analysis:Reading scores are considered in the screening or selection index. Medical schools may well be cautious of how heavily they weight Physics scores in any selection index, since this may disadvantage females in a way that cannot be justified. Studies such as the present one attest to the importance of considering these and other factors in determining the way in which MCAT scores are used in making admissions decisions.

BEST COPY

REFERENCES

Association of American Medical Colleges. 1983 MCAT Summary of Score Distributions. Washington: AAMC, 1983.

Cleary, T.A. Test bias: prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Gulliksen, H., and Wilks, S.S. Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

Linn, R.L. Ability testing; individual differences, prediction and differential prediction. In Ability testing: uses, consequences, and controversies. A Wigdor and W. Garner (Eds.). Washington, D.C.: National Academy Press, 1982, 335-388.

Linn, R.L., and Werts, C. Considerations for studies of test bias. Journal of Educational Measurement, 1971, 8, 1-4.

Maccoby, E., and Jacklin, C. The psychology of sex differences. Stanford, California: Stanford University Press, 1975.

Wang, M., and Wilson-Pessano, S. Report on the Investigation of Item Bias in the New MCAT. Unpublished paper. Palo Alto: American Institutes for Research, 1982.

BEST COPY